TITLE: Non-Contact Thermometers for Detecting Fever: A Review of Clinical Effectiveness

DATE: 20 November 2014

CONTEXT AND POLICY ISSUES

Body temperature can be measured in a number of ways. Traditionally, body temperature has been measured using contact thermometers that are placed on the forehead or in the mouth, ear, armpit or rectum.[1] For children in particular, rectal temperature measurement is often considered to be the gold standard.[2] Non-contact thermometers allow a person's temperature to be taken with minimal (tympanic) or no (Non-contact infrared thermometer [NCIT], thermal scanner) contact with the person. This means temperature can be measured without the discomfort of having to sit still with a thermometer in the mouth, armpit, or rectum long enough to obtain a correct temperature reading.[1] The lack of contact also means the disinfection process between patients for the thermometers is minimal or unnecessary, allowing for easier and faster use when screening large numbers of people in settings like airports or border crossings.[1]

The main types of non-contact thermometers are non-contact infrared thermometers, tympanic thermometers, and thermal scanners. Non-contact infrared thermometers are held three to 15 cm away from the patient and typically measure temperature on the forehead[1] or temple.[3] Tympanic thermometers measure the thermal radiation from the tympanic membrane and within the ear canal.[2] Handheld thermal scanners can be used to take a person's temperature from a greater distance than other non-contact thermometers, which may make them a good candidate for use in mass screening situations.[1] The optimal cut-off temperature for determining fever differs for each device.[1] However, not everyone who has an infection or is infectious will have a fever. Additionally, fevers can be lowered by using antipyretic medications.[1]

The objective of this report is to determine the effectiveness and accuracy of non-contact thermometers for the detection of febrile individuals.

RESEARCH QUESTIONS

1. What is the accuracy of tympanic infrared thermometers for detecting febrile individuals?

2. What is the accuracy of handheld infrared non-contact thermometers for detecting febrile individuals?

3. What is the accuracy of thermal scanners for detecting febrile individuals?

4. What is the comparative effectiveness of tympanic thermometers, handheld infrared thermometers, and thermal scanners for detecting febrile individuals?

KEY FINDINGS

Evidence retrieved from sixteen non-randomized studies and four systematic reviews (SRs) supports the accuracy of tympanic thermometers and, more cautiously, of thermal scanners. Evidence for the accuracy of infrared skin thermometers is equivocal and requires more research. However, the generalizability of the evidence found is questionable.

METHODS

Literature Search Strategy

A limited literature search was conducted on key resources including PubMed, The Cochrane Library (2014, Issue 10), University of York Centre for Reviews and Dissemination (CRD) databases, Canadian and major international health technology agencies, as well as a focused Internet search. No filters were applied to limit the retrieval by study type. Where possible, retrieval was limited to the human population. The search was also limited to English language documents published between January 1, 2009 and 15 October, 2014.

Selection Criteria and Methods

One reviewer screened citations and another reviewer selected studies based on full-text review. In the first level of screening, titles and abstracts were reviewed and potentially relevant articles were retrieved and assessed for inclusion. The final selection of full-text articles was based on the inclusion criteria presented in Table 1.

Table 1: Selection Criteria

Population	Any
Intervention	Tympanic thermometers, handheld infrared thermometers, thermal scanners
Comparator	Devices compared to each other or to a reference standard
Outcomes	Diagnostic accuracy (true/false positives/negatives, agreement with reference standard)
Study Designs	Health technology assessments, systematic reviews, meta-analyses, randomized controlled trials, non-randomized studies

Exclusion Criteria

Studies were excluded if they did not satisfy the selection criteria, if they were duplicate publications, if they were included in a selected SR, if they were non-systematic reviews or were published prior to 2009.

Critical Appraisal of Individual Studies

Systematic reviews were appraised using the AMSTAR (A Measurement Tool to Assess Systematic Reviews) checklist.[4] Items included in the AMSTAR checklist include a priori design of the review, eligibility criteria, information sources searched, study selection, data items and methods of data extraction, quality of studies, interpretation of the results, publication bias, and sources of funding.[4]

Comparative non-randomized controlled trials were appraised using the Downs and Black checklist. Items evaluated included clear study objectives, clear study inclusion and exclusion criteria, clear description of potential confounders, description of losses to follow up, blinding, appropriate statistical tests used., accuracy of the outcome measures, and whether power was sufficient to detect a difference if one existed.[5] A numeric score was not calculated. Strengths and limitations were reviewed for included studies.

SUMMARY OF EVIDENCE

Quantity of Research Available

A total of 523 citations were identified in the literature search. Following screening of titles and abstracts, 498 citations were excluded and 25 potentially relevant reports from the electronic search were retrieved for full-text review. Five potentially relevant publications, of a total of 14, were retrieved from the grey literature search. Of these 30 potentially relevant articles, 10 publications were excluded for various reasons, while 20 publications met the inclusion criteria and were included in this report. Of the studies included, four are systematic reviews and sixteen are non-randomized studies. Appendix 1 describes the PRISMA flowchart of the study selection.

The summary of study characteristics table is provided in Appendix 2, the results of the critical appraisal are in Appendix 3, and the main study findings and author conclusions are provided in Appendix 4.

Summary of Study Characteristics

<u>The accuracy of tympanic thermometers for detecting febrile individuals</u>

A total of fifteen studies were identified that evaluated the accuracy of tympanic thermometers. Of these publications, four were systematic reviews[6-9] and eleven were non-randomized studies.[3,10-19] Among these publications, tympanic temperature was used as the reference in four studies[6,10,11,13] where results did not focus on the tympanic measurements.

Country of origin

Systematic reviews originated from China,[8,9] New-Zealand,[7] and France.[6] The non-randomized studies were from Korea,[10] China,[11] Thailand,[12] New-Zealand,[13] USA,[14] Spain,[3] Gabon,[15] United-Kingdom,[17] Pakistan,[18] Malaysia,[16] and Belgium.[19]

Population

The mean age of patients in the included studies ranged from neonates[9] to 80.9[19] years and a majority of studies reported a ratio of male/female near 1:1. Some studies only included pediatric patients,[8,9,15] whereas one study only included geriatric patients.[19] Studies included inpatients or patients presenting at hospital,[3,6-9,11,14-19] or travelers presenting at borders.[10,12,13] The sample size of the non-randomized studies ranged from 21[16] to 2000.[18] The systematic reviews included from 3[7] to 31[8] studies with samples from 9[7] to 72,327[6] participants.

Interventions and comparators

Devices used to measure tympanic temperature varied across studies. The BraunThermoScan and the FirstTemp Genius were those used most often, whereas one study did not report the model of the device.[11] The number of measurements and the mode of the device (i.e. the algorithm transforming the actual reading into the predicted body temperature) may have been different between studies, but were not always specified.

The accuracy of tympanic thermometers was compared with rectal temperature in six studies,[8,9,14-16,19] with oral temperature in two studies,[12,18] with pulmonary artery catheter temperature in two studies,[3,7] with axillary temperature in one study,[15] and with nasopharyngeal temperature in one study.[17]

Years of publication

The years of publication ranged from 2009 to 2014.

<u>The accuracy of handheld infrared skin thermometers for detecting febrile individuals</u>

Seven studies evaluated the accuracy of handheld infrared skin thermometers. Of these publications, one was a systematic review[6] and six were non-randomized studies.[3,17,20-23]

Country of origin

The systematic review originated from France.[6] The non-randomized studies were from Bolivia,[23] Italy,[20] Spain,[3] USA,[21,22] and United-Kingdom.[17]

Population

The age of the patients included in the studies ranged from 1 month[21] to over 80[17] years and most of the studies reported a similar proportion of males and females. Some studies only included pediatric patients.[20-23] All studies included inpatients or patients presenting at hospital. The sample size of the non-randomized studies ranged from 61[17] to 855.[22] The six studies included in the systematic reviews have samples ranging from 176 to 72,327 participants.[6]

Interventions and comparators

Devices used to measure skin temperature varied across studies. The Thermofocus and the Exergen thermometers were those used in the non-randomized studies, whereas one study did not report the model of the device[3] and the SR included studies with other kinds of devices.[6] The number of measurements and the mode of the device (i.e. the algorithm transforming the actual reading into the predicted body temperature) were different or not reported between studies.

The accuracy of handheld infrared skin thermometers was compared with rectal temperature in two studies,[21,23] with pulmonary artery catheter temperature in one study,[3] with tympanic thermometers in one study,[6] with axillary temperature in one study,[20] with nasopharyngeal temperature in one study,[17] and with a reference that could be either oral, rectal or axillary temperatures in one study.[22]

Years of publication

The years of publication ranged from 2009 to 2013.

<u>The accuracy of thermal scanners (infrared cameras) for detecting febrile individuals</u>

Six studies evaluated the accuracy of thermal scanners. Of these publications, one was a systematic review[6] and five were non-randomized studies.[10,11,13,22,24]

Country of origin

The systematic review originated from France.[6] The non-randomized studies were from Korea,[10] China,[11] New-Zealand,[13] and USA.[22,24]

Population

The age of the patients included in the studies ranged from 6 months[22] to 92[24] years and most of the studies reported a similar proportion of males and females. One study only included pediatric patients.[22] Studies included inpatients or patients presenting at hospital,[6,11,22,24] or travelers presenting at borders.[10,13] The sample size of the non-randomized studies ranged from 608[10] to 2873.[24] The studies included in the systematic reviews have samples ranging from 176 to 72,327 participants.[6]

Interventions and comparators

Devices used to measure skin temperature varied across studies. The FLIR and the OptoTherm ThermoScreen were those used in the majority of studies.

The accuracy of thermal scanners was compared with tympanic thermometers in four studies,[6,10,11,13] with oral temperature in one study[24] and with a reference that could be either oral, rectal or axillary temperatures in one study.[22]

Years of publication

The years of publication ranged from 2009 to 2014.

Summary of Critical Appraisal

The SRs[7-9] had a pre-specified protocol, except Bitar et al.[6] which did not describe an a priori study design. Two[8,9] SRs had study selection performed in duplicate by independent reviewers. The literature search strategy, including grey literature search, was described in two SRs,[8,9] whereas Jefferies et al. did not mention a grey literature search[7] and Bitar et al. did not perform a comprehensive literature search (only one database search, no mention of grey literature).[6] Excluded studies were not disclosed in any of the SRs. A list of the included studies, with their characteristics and an assessment of their individual quality were reported in two SRs,[7,8] whereas Zhen et al.[9] and Bitar et al.[6] did not report individual quality. The conclusions of one SR were in line with the quality of its results,[7] whereas Zhen et al. expressed different conclusions that were in contradiction with each other.[8] However, it was the only SR[8] where publication bias has been assessed. The heterogeneity and the comparability of the data was assessed in the three SRs[7-9] where a meta-analysis was planned, even though it could not be performed in on case.[7] Conflicts of interest of the included studies were not documented in any of the SRs. Other issues were: subjective inclusion criteria,[9] lack of information on the included studies,[7] poor statistical methods in the included studies,[9] and the non-blinded design of the included studies.[6-9]

The common strengths of the non-randomized studies were the objectivity of the measurements (i.e. body temperature) and the comparison in a single group.The common limitation of the non-randomized studies is that they all used, at least in part, non-blinded investigators for the assessment of temperatures. Some studies also lacked a sample size calculation,[10,11,14,18,19,21,23] a description of the percentage of eligible participants who were enrolled,[3,11,14-18,20,21,23] a description of the thermometers used,[3,11] statistical tests or P-values,[13,14,16] or a description of statistical analyses.[11] Moreover, a few studies based their conclusions on very small numbers of feverish subjects[10,13,18] or failed to clearly state the outcomes in the introduction.[12,23] Cho et al. conducted a retrospective study and the variability of measurements have not been correctly reported.[10] Oyakhirome et al. did not use the gold standard comparator (rectal temperature) in all patients and did not state explicit inclusion/exclusion criteria.[15] Since Priest et al. only analyzed a small proportion of eligible participants the representativity of their sample is questionable.[13] Participant characteristics were not well desribed in both Oyakhirome and Priest studies.[13,15] Other issues encountered were conclusions that could not be extrapolated to febrile patients[17] and a study where some patients of interest have been excluded.[12]

The time elapsed between the different measurements was short or nonexistent in twelve studies,[3,12,14-21,23,24] but has not been specifically reported in four studies.[10,11,13,22]. The choices for the reference temperature were reasonable in most of the studies as the utilization of the actual core temperature (pulmonary artery catheter) or the common gold standard (rectal temperature) is not always feasible, depending of the context of the study. Nevertheless, the studies which compared to these two standards are more likely to have an accurate reference for the core body temperature.

Summary of Findings

<u>The accuracy of tympanic thermometers for detecting febrile individuals</u>

The systematic review of pediatric studies from Zhen et al.[8] reported a sensitivity of 0.70 (95% confidence interval [CI] 0.68 to 0.72), a specificity of 0.86 (95% CI 0.85 to 0.88), a positive likelihood ratio of 9.14 (95% CI 6.37 to 13.11), a negative likelihood ratio of 0.24 (95% CI 0.17 to

0.34), a diagnostic odds ratio of 47.3 (95% CI 29.79 to 75.18) and an area under the receiver operating characteristic curve (AUROC) of 0.94 when comparing tympanic thermometry with rectal thermometry. The overall pooled mean difference between tympanic and rectal temperature was 0.22ºC (95% limits of agreements [LOA] -0.44 to 1.30ºC).[9] This difference was reduced to 0.15ºC (95% LOA -0.32 to 1.10ºC) when considering a subgroup of febrile children.[9] Compared to pulmonary artery catheter temperature, the SR from Jefferies et al. reported a mean difference from tympamic temperature within the ±0.2 ºC range.[7] Two SRs expressed conclusions in favor of the utilization of tympanic thermometry, [7,8] whereas one stated that its accuracy (with an LOA spanning over 1.74 ºC) is poor.[9]

Compared with oral temperature, Chue et al reported a mean difference ranging from 0.05ºC (95% CI 0.01 to 0.08) to 0.12ºC (95% CI 0.07 to 0.17) depending of the investigator.[12] Rabanni et al. reported a mean difference of 0.1ºC, a correlation of 0.723, a sensitivity of 66%, a specificity of 99.6%, a positive predicitive value (PPV) of 91%, and a negative predictive value (NPV) 98%.[18]

When comparing with rectal temperature, Barnett et al. reported a mean difference of 0.22ºF (95% CI -1.61 to 2.05),[14] Smitz et al. found 95% LOA of -0.83 to 0.42ºC for ThermoScan and -1.32 to 0.20ºC for Genius[19] and Oyakhirome reported mean difference of 0.3ºC (95% CI 0.2 to 0.3, LOA -1 to 2).[15] Across studies, reported sensitivities were of 63.6%,[16] 74.12%,[14] 75%[15] and 94%;[19] reported specificities were of 86.22%,[14] 95%,[15] 97.4%[16] and 96-98%;[19] reported PPV were of 55.26%,[14] 87.5%,[19] 85-89%[16] and 94%;[15] reported NPV were 76%,[15] 90.5%,[16] 93.55%[14] and 99%;[19] reported correlation coefficients were 0.806[16] and 0.84-0.91[19]. Oyakhirome reported an optimal fever cutoff point of 100.2ºF with an AUROC of 0.878.[14]

Compared to pulmonary artery catheter temperature at a threshold of 38.5ºC,[3] tympanic thermometry had a specificity of 98%, a PPV of 89%, a NPV of 98% and an AUROC of 0.987±0.007.[3] Sensitivity was not reported. Compared to nasopharyngeal probe, tympanic thermometry had a mean difference of 0.19ºC (95% LOA -0.32 to 0.71) or 0.98ºC (95% LOA 0.42 to 1.54) depending on the device.[17]

Six studies expressed conclusions in favor of the utilization of tympanic thermometry,[3,12,16-19] whereas one study stated that the variability of measurements with tympanic termometry was too high.[14] One study did not express conclusions in favor or against use of the device.[15]

<u>The accuracy of handheld infrared skin thermometers for detecting febrile individuals</u>

The systematic review from Bitar et al.[6] reported sensitivities ranging from 4.0 to 89.6%, specificities ranging from 75.4 to 99.6%, positive likelihood ratios ranging from 0.9 to 76.0%, negative likelihood ratios ranging from 86.1 to 99.7%, correlation coefficients ranging from 0.25 to 0.71, and AUROC ranging from 0.86 to 0.96 when comparing infrared non-contact thermometers (including both skin thermometers and cameras) with tympanic thermometry. The authors of this SR highlighted the poor scientific evidence available for the utilization of infrared skin thermometers and thermal scanners for mass screening.[6]

Across studies, comparators were rectal temperature,[21,23] axillary temperature,[20] pulmonary artery catheter temperature,[3] nasopharyngeal probe temperature,[17] or either oral, rectal or axillary temperature.[22] Teran et al. reported a mean difference of 0.029 ± 0.01ºC and of -0.02 ± 0.277ºC depending of the model used.[23] Chiappini et al. reported a mean difference of 0.11ºC.[20] Fortuna et al. reported a mean difference of -0.1ºF.[21] Mangat et al. reported a mean difference

of 0.66ºC (95% LOA -0.15 to 1.48).[17] Reported sensitivities were 76.8%,[22] 89%[20] and 91-97%;[23] reported specificities were of 79.4%,[22] 83%,[3] 90%[20] and 97-99.6%;[23] reported PPV were of 47%,[3] 70%[20] and 95.2-99.3%;[23] reported NPV were of 96%,[3] 97%[20] and 94.6-98.1%;[23] reported correlation coefficients were of 0.48,[21] 0.66,[22] 0.837[20] and 0.950-0.952;[23] reported AUROC were of 0.852[22] and 0.853 ± 0.051.[3]

Three studies expressed conclusions in favor of the utilization of infrared skin thermometry,[20,22,23] whereas three studies stated that this type of device is lacking accuracy.[3,17,21]

The accuracy of thermal scanners (infrared cameras) for detecting febrile individuals

The findings of a systematic review[6] that included studies both with infrared skin thermometers and thermal scanners have been described in the previous sub-section.

Nguyen et al. compared thermal scanners with oral thermometers.[24] Selent et al. compared with oral, rectal or axillary temperature.[22] All other studies compared with tympanic temperature[9,13] or tympanic + oral temperatures.[11] Cho et al. reported a mean difference of -1.31ºC but this was not statistically different from tympanic temperature (P = 0.316)[10] whereas Chan et al. reported a mean difference of -3.10ºC.[11] Reported sensitivities were 87%,[11] 83.0-83.7%,[22] 86%[13] and 80.0-91.0%;[24] reported specificities were 34-43%,[11] 85.7-86.3%,[22] 71%[13] and 65.0-86.0;[24] reported PPV were 10-11%,[11] 1.5%[13] and 5.7-18.4%;[24] reported NPV were 97-98%[11] and 99.1-99.6%;[24] a positive likelihood of 1.33-1.53[11] was reported; a negative likelihood of 0.29-0.37[11] was reported; reported correlation coefficients were < 0.5[11] and 0.75-0.78;[22] reported AUROC were 0.780-0.815,[11] 0.922-0.923%[22] 0.86[13] and 0.78-0.96.[24]

Four studies expressed conclusions in favor of the utilization of thermal scanners for fever detection,[10,13,22,24] whereas one study stated that this type of device is unsuitable for this purpose due to a high proportion of false positives.[11]

Limitations

The most common limitation across studies is that they all used, at least in part, non-blinded investigators for the assessment of temperature, but given the objective nature of temperature measurement, this should not be considered a major biasing limitation. Moreover, for many studies it is not clear if they were powered to find a difference between their devices.[10,11,14,18,19,21,23] Also, many studies failed to reveal the percentage of eligible participants who were actually enrolled.[3,11,14-18,20,21,23] This is of importance since it is not clear if the samples were representative of the population. The profile of people who refused to participate to the studies has not been described. Therefore, it is plausible that feverish or very ill people might be underestimated in those studies.

Across studies, many potential confounders of body temperature have been mentioned such as sweat, gender, age, the range of temperature, the rater, physical activity, the use of antipyretic drugs and the emotional state, but the list is not exhaustive. It has to be kept in mind that those factors can bias the results of the study reviewed, especially when using non-contact infrared (including tympanic, skin or scanners) thermometers. Many studies were specifically conducted on a pediatric population and one was conducted on an geriatric population. Since age is a potential counfounder,[11,18,22] the generalizability of those studies to the adult population is questionable.

As mentioned by Zhen et al.,[8,9] there is a lot of heterogeneity in the data between studies. Some specific factors affect the comparability of the studies. Reviewed studies have been using different thermometric devices that, depending of the brand, model and mode used, convert the actual reading to a different output measure following their own algorithm. Also, threshold temperatures for fever varied across studies. Some studies aimed to find the optimal threshold for their device even if it was different by many degrees from the reference.[11]

CONCLUSIONS AND IMPLICATIONS FOR DECISION OR POLICY MAKING

Most of the non-randomized studies included in this review had a similar prospective observationnal design with non-blinded measurements taken in a single group. Two SRs were deemed of average quality and two had many limitations. A majority of studies used rectal, oral, axillary, tympanic, or pulmonary artery catheter as a reference for body temperature. Seven out of twenty publications were specifically investigating pediatric patients, while only one enrolled geriatric patients. Three studies were conducted at a border crossing; others were in hospitals. The most commonly reported outcomes were sensitivity, specificity, PPV, NPV, positive/negative likelihood ratios, corelation coefficient and AUROC.

The conclusions of six non-randomized studies and two SRs supported the utilization of tympanic thermometry. The conclusions from one study and one SR were not in favor of its accuracy. The evidence is then in favor of the accuracy of tympanic thermometers. The accuracy of handheld infrared skin thermometers were favored by three studies but also unfavored by three studies. Four studies expressed conclusions in favor of the utilization of thermal scanners for fever detection, whereas one study stated that this type of device is unsuitable for this purpose. The conclusions of a SR, although of low quality, highlighted the poor scientific evidence available for the utilization of infrared skin thermometers and thermal scanners for mass screening. Evidence for the accuracy of infrared skin thermometers is equivocal whereas it is somehow in favor of the accuracy of thermal scanners.

Many issues raise doubts about the generalizability of the included studies. It is not clear if the people who refused to participate in these studies biased the results and the percentage of enrollment among eligible participants was not reported in most of the studies. The retrieved studies have mentioned potential confounders for measure of temperature such as sweat, gender, age, the range of temperature, the rater, physical activity, the use of antipyretic drugs and emotional state. These factors are even more susceptible to vary in a real world conditions than in a clinical study setting. Moreover, the different brand/model/mode of devices used make it difficult to draw general conclusions on a class of thermometers. Also, a fair number of pediatric studies were included in the present review, limiting the extrapolation of their results to a general population.

Depending on the context of utilization (hospital vs border), the volume of measurements to be done and the age of the person to be measured, it might be imperative to use infrared thermometers over more accurate and/or more invasive thermometers. Therefore, tympanic thermometers and thermal scanners might be the only effective and accurate tools to detect fever under certain circumstances. However, one has to keep in mind that screening for fever and screening for a virus are two different issues.

In conclusion, evidence retrieved from sixteen non-randomized studies and four systematic reviews is in favor of accuracy of tympanic thermometers and, more cautiously, of thermal

scanners. Evidence for the accuracy of infrared skin thermometers is equivocal and requires more research. The generalizability of the evidence found is nevertheless uncertain.

PREPARED BY:
Canadian Agency for Drugs and Technologies in Health
Tel: 1-866-898-8439
www.cadth.ca

REFERENCES

1. Non-contact temperature measurement devices: considerations for use in port of entry screening activities [Internet]. Washington (DC): Centers for Disease Control and Prevention; 2014. [cited 2014 Oct 23]. Available from: http://wwwnc.cdc.gov/travel/pdf/ebola-non-contact-temperature-measurement-guidance.pdf

2. Leduc C, Woods S, Canadaian Paediatric Society Community Paediatrics Committee. Position statement: temperature measurement in paediatrics [Internet]. Ottawa: Canadian Paediatric Society; 2013. [cited 2014 Oct 23]. Available from: http://www.cps.ca/en/documents/position/temperature-measurement

3. Rubia-Rubia J, Arias A, Sierra A, Aguirre-Jaime A. Measurement of body temperature in adult patients: comparative study of accuracy, reliability and validity of different devices. Int J Nurs Stud. 2011 Jul;48(7):872-80.

4. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol [Internet]. 2007 [cited 2014 Nov 17];7:10. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1810543/pdf/1471-2288-7-10.pdf

5. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health [Internet]. 1998 Jun [cited 2014 Nov 17];52(6):377-84. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1756728/pdf/v052p00377.pdf

6. Bitar D, Goubar A, Desenclos JC. International travels and fever screening during epidemics: a literature review on the effectiveness and potential use of non-contact infrared thermometers. Euro Surveill [Internet]. 2009 Feb 12 [cited 2014 Oct 24];14(6). Available from: http://www.eurosurveillance.org/images/dynamic/EE/V14N06/art19115.pdf

7. Jefferies S, Weatherall M, Young P, Beasley R. A systematic review of the accuracy of peripheral thermometry in estimating core temperatures among febrile critically ill patients. Crit Care Resusc. 2011 Sep;13(3):194-9.

8. Zhen C, Xia Z, Ya JZ, Long L, Jian S, Gui JC, et al. Accuracy of infrared tympanic thermometry used in the diagnosis of fever in children: a systematic review and meta-analysis. Clin Pediatr (Phila). 2014 Aug 6.

9. Zhen C, Xia Z, Long L, Pu Y. Accuracy of infrared ear thermometry in children: a meta-analysis and systematic review. Clin Pediatr (Phila). 2014 Oct;53(12):1158-65.

10. Cho KS, Yoon J. Fever screening and detection of febrile arrivals at an international airport in Korea: association among self-reported fever, infrared thermal camera scanning, and tympanic temperature. Epidemiol Health [Internet]. 2014 [cited 2014 Nov 17];36:e2014004. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4101989/pdf/epih-36-e2014004.pdf

11. Chan LS, Lo JL, Kumana CR, Cheung BM. Utility of infrared thermography for screening febrile subjects. Hong Kong Med J [Internet]. 2013 Apr [cited 2014 Oct 24];19(2):109-15. Available from: http://www.hkmj.org/article_pdfs/hkm1304p109.pdf

12. Chue AL, Moore RL, Cavey A, Ashley EA, Stepniewska K, Nosten F, et al. Comparability of tympanic and oral mercury thermometers at high ambient temperatures. BMC Res Notes [Internet]. 2012 [cited 2014 Oct 24];5:356. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3496650/pdf/1756-0500-5-356.pdf

13. Priest PC, Duncan AR, Jennings LC, Baker MG. Thermal image scanning for influenza border screening: results of an airport screening study. PLoS ONE [Internet]. 2011 [cited 2014 Oct 24];6(1):e14490. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3016318/pdf/pone.0014490.pdf

14. Barnett BJ, Nunberg S, Tai J, Lesser ML, Fridman V, Nichols P, et al. Oral and tympanic membrane temperatures are inaccurate to identify fever in emergency department adults. West J Emerg Med [Internet]. 2011 Nov [cited 2014 Oct 24];12(4):505-11. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3236177/pdf/wjem-12-04-505.pdf

15. Oyakhirome S, Profanter K, Kremsner PG. Assessment of fever in African children: implication for malaria trials. Am J Trop Med Hyg [Internet]. 2010 Feb [cited 2014 Oct 24];82(2):215-8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2813159/pdf/tropmed-82-215.pdf

16. Dzarr AA, Kamal M, Baba AA. A comparison between infrared tympanic thermometry, oral and axilla with rectal thermometry in neutropenic adults. Eur J Oncol Nurs. 2009 Sep;13(4):250-4.

17. Mangat J, Standley T, Prevost A, Vasconcelos J, White P. A comparison of technologies used for estimation of body temperature. Physiol Meas. 2010 Sep;31(9):1105-18.

18. Rabbani MZ, Amir M, Malik M, Mufti M, Bin PM, Iftekhar S. Tympanic temperature comparison with oral mercury thermometer readings in an OPD setting. J Coll Physicians Surg Pak. 2010 Jan;20(1):33-6.

19. Smitz S, Van de WA, Smitz MF. Reliability of infrared ear thermometry in the prediction of rectal temperature in older inpatients. J Clin Nurs. 2009 Feb;18(3):451-6.

20. Chiappini E, Sollai S, Longhi R, Morandini L, Laghi A, Osio CE, et al. Performance of non-contact infrared thermometer for detecting febrile children in hospital and ambulatory settings. J Clin Nurs. 2011 May;20(9-10):1311-8.

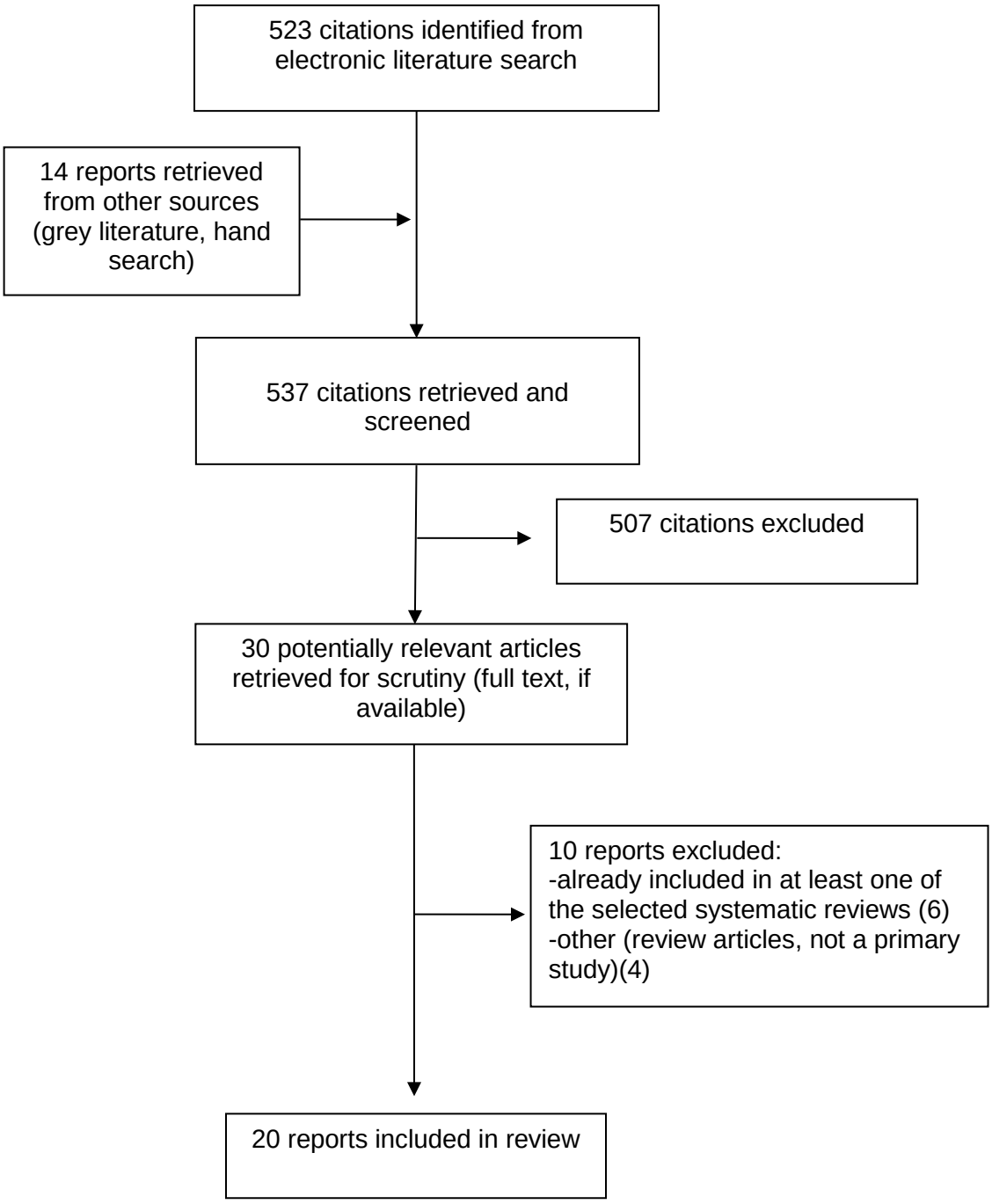
21. Fortuna EL, Carney MM, Macy M, Stanley RM, Younger JG, Bradin SA. Accuracy of non-contact infrared thermometry versus rectal thermometry in young children evaluated in the emergency department for fever. J Emerg Nurs. 2010 Mar;36(2):101-4.

22. Selent MU, Molinari NM, Baxter A, Nguyen AV, Siegelson H, Brown CM, et al. Mass screening for fever in children: a comparison of 3 infrared thermal detection systems. Pediatr Emerg Care. 2013 Mar;29(3):305-13.

23. Teran CG, Torrez-Llanos J, Teran-Miranda TE, Balderrama C, Shah NS, Villarroel P. Clinical accuracy of a non-contact infrared skin thermometer in paediatric practice. Child Care Health Dev. 2012 Jul;38(4):471-6.

24. Nguyen AV, Cohen NJ, Lipman H, Brown CM, Molinari NA, Jackson WL, et al. Comparison of 3 infrared thermal detection systems and self-report for mass fever screening. Emerg Infect Dis [Internet]. 2010 Nov [cited 2014 Oct 24];16(11):1710-7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3294528/pdf/10-0703_finalR.pdf

APPENDIX 1: Selection of Included Studies



APPENDIX 2: Summary of Included Studies

First Author, Publication Year, Country	Study design	Patients Characteristics, Sample Size (n)	Intervention	Comparator(s)	Outcomes
Systematic Reviews					
Zhen, 2014,[8] China	SR/MA of studies (29/31 prospective) on the diagnosis of pediatric fever.	31 studies (25 articles), 5749 pediatric patients. Sample size range from 40 to 964 patients. Age: < 18 years.	Infrared tympanic thermometry	Rectal thermometry (electronic or mercury)	▪ True/false positives/negatives ▪ Sensitivity ▪ Specificity ▪ Positive/negative predictive value
Zhen, 2014,[9] China	SR/MA of cross-sectional, prospective studies investigating thermometry in pediatric patients.	28 studies (33 comparisons), 5448 pediatric patients. Sample size range from 36 to 623 patients. Age: < 16 years.	Infrared ear thermometry	Rectal thermometry (electronic or mercury)	▪ Mean difference from comparator ▪ Upper and lower 95% limits of agreement
Jefferies, 2011,[7] New Zealand	SR of prospective studies investigating thermometry in critically ill patients.	3 studies, 110 critically ill adult patients with fever. Sample size ranging from 9 to 72 patients.	Infrared tympanic thermometry. Studies compared different devices/modes/ core temperature ranges	Pulmonary artery catheter core thermometry	▪ Mean difference from core temp.
Bitar, 2009,[6] France	SR of studies on fever screening under mass screening conditions	6 studies, 77,024 participants (including healthy visitors, hospitalized patients or patients presenting for emergency or consultation). Sample size ranging from 176 to 72,327.	Non-contact thermometry: infrared skin thermometers and thermal infrared cameras (tympanic was considered contact)	Tympanic thermometry	▪ Sensitivity ▪ Specificity ▪ Positive/negative predictive values

First Author, Publication Year, Country	Study design	Patients Characteristics, Sample Size (n)	Intervention	Comparator(s)	Outcomes
Non-randomized studies					
Cho, 2014,[10] Korea	Non-blinded, retrospective, cross-sectionnal study. Airport setting.	608 symptomatic arrivals (runny/stuffed nose, sore throat, cough, fever) were analysed. 313 F /294 M. Age: mean 25.1 years, range 1 to 86. Participants were travelers at an international airport.	Thermal camera temperature (Thermovision A20M, ThermoTracer TH7800, ThermoGraphy R300)	Tympanic (or ear) thermometry (ThermoScan IRT-3020 and IRT-4020)	▪ Fever prevalence ▪ Association between self-reported fever and tympanic temperature ▪ Difference between thermal camera and tympanic thermometry
Chan, 2013,[11] China	Non-blinded, prospective study. Hospital setting.	1517 patients with or without fever, who presented to an accident and emergency department. Mean age: 45.8 years. 747 M/770 F	Remote sensing infrared camera (maximal frontal, lateral views and forehead temperatures) (FLIR ThermaCAM S40)	Oral or ear thermometry	▪ Proportion of feverish subjects detected ▪ Correlation with reference ▪ Sensitivity ▪ Specificity ▪ PPV/NPV ▪ Positive/negative likelihoods
Selent, 2013,[22] USA	Non-blinded, prospective study. Hospital setting.	855 pediatric patients who presented at emergency department. 469 boys/386 girls. Age: 6 months to 17 years. 218 rectal, 422 oral and 215 axillary temperature.	3 ITDS: two thermal cameras (OptoTherm Thermoscreen and FLIR) and one handheld infrared skin thermometer (Thermofocus)	Oral, rectal or axillary thermometry following age.	▪ Sensitivity ▪ Specificity ▪ Correlation with reference ▪ Receiver operating characteristic curve

First Author, Publication Year, Country	Study design	Patients Characteristics, Sample Size (n)	Intervention	Comparator(s)	Outcomes
Chue, 2012,[12] Thailand	Non-blinded, prospective study. Screening at border with high ambiant temperature.	201 persons who presented at a border. Age: mean 27 years, range 5 to 70. 40.8% M/59.2% F. 26.9% pregnant women.	Tympanic thermometry (Braun ThermoScan IRT 4520)	Oral mercury in glass thermometry	▪ Temperature difference from oral thermometry
Teran, 2012,[23] Bolivia	Non-blinded, prospective study. Hospital setting.	434 pediatric patients at emergy room or as inpatient. Age 1 to 48 months. Mean 14.6 ±10.7 months. 208 males/ 226 females.	Infrared non-contact skin (Thermofocus) thermometry and temporal artery (Exergen) thermometry	Rectal glass mercury thermometer	▪ Temperature difference from comparators ▪ Correlation vs comparators ▪ Sensitivity ▪ Specificity ▪ Positive predictive value ▪ Negative predictive value
Chiappini, 2011,[20] Italy	Non-blinded, prospective multicenter (hospital) study.	251 pediatric patients with stable, non-chronic, conditions admitted for any reason. Age: median 4.5 years, range from 1 month to 18 years. 50.6% M/49.4% F	Non-contact infrared thermometer (Thermofocus, mid-forehead temperatures)	Axillary temperature measurement with mercury thermometers	▪ Variability of repeated measures ▪ Concordance between forehead and axillary measures ▪ Discomfort assessment ▪ sensitivity ▪ specificity ▪ PPV ▪ NPV

First Author, Publication Year, Country	Study design	Patients Characteristics, Sample Size (n)	Intervention	Comparator(s)	Outcomes
Priest 2011,[13] New Zealand	Non-blinded prospective observational study. Airport setting.	1275 airline travellers during a seasonal influenza outbreak. (1275 in temperature comparison, 1268 for influenza prediction) All symptomatic travellers were invited to have throat and nose swabs and temperature measurement. Other travellers were randomly asked to participate.	Infrared thermal image scanners (FLIR) to measure cutaneous temperature	Tympanic temperature measurement (ThermoScan) and respiratory sampling	▪ Receiver operating characteristic curve sensitivity ▪ specificity ▪ PPV
Barnett, 2011,[14] USA	Non-blinded prospective study. Hospital setting.	457 patients in emergency department. Average age = 64 years (SD 19 years), range 18 to 96. 59% Females/41% Males.	Tympanic membrane (First Temp Genius II) and oral (IVAC Temp Plus II) thermometry	Rectal (IVAC Temp Plus II) thermometry	▪ Difference from reference ▪ Sensitivity ▪ Specificity ▪ Predictive values (PPV, NPV) ▪ Likelihood ratios ▪ ROC curves
Rubia-Rubia 2010,[3] Spain	Non-blinded prospective study. Hospital setting.	201 adult patients from intensive care unit. Mean age 59 (SD 10) years. 74% M/26% F.	Infrared ear and frontal thermometers Gallium-in-glass, reactive strip, and digital in axilla All compared to core temperature	Core body temperature measured at the pulmonary artery	▪ Validity ▪ Reliability ▪ Accuracy ▪ External Influence ▪ Waste Generated ▪ Ease Of Use ▪ Speed ▪ Durability ▪ Security ▪ Comfort

First Author, Publication Year, Country	Study design	Patients Characteristics, Sample Size (n)	Intervention	Comparator(s)	Outcomes
			simultaneously		▪ Cost
Fortuna, 2010,[21] USA	Non-blinded prospective observational study. Hospital setting.	Convenience sample of 200 children from 1 month to 4 years of age presenting to tertiary pediatric emergency department.	Non-contact infrared thermometer (Thermofocus) (mid-forehead)	Rectal thermometer (Welch Allen SureTemp)	▪ Agreement in measurement between two techniques ▪ Bias of techniques
Nguyen, 2010,[24] USA	Non-blinded prospective study. Hospital setting.	2,873 adults (≥18 years of age) presenting to hospital emergency departments. 52.7% M/47.3% F. Age: mean 42, range 18 to 92.	3 ITDS (cameras): OptoTherm, FLIR, and Wahl	Oral digital thermometry	▪ Sensitivity ▪ Specificity ▪ Receiver operating characteristic curve ▪ PPV/NPV ▪ Accuracy compared to oral thermometry
Oyakhirome, 2010,[15] Gabon	Non-blinded prospective study. Hospital setting.	1,000 children aged ≤10 years presenting to hosptial outpatient department with complaint of fever. Rectal measurements for 835.	Tympanic thermometry (Braun ThermoScan 6022)	Rectal (“gold standard”) and axillary thermometry (Thermoval Basic)	▪ Sensitivity ▪ Specificity ▪ PPV/NPV ▪ Spearman rank correlation coefficients for paired readings ▪ Mean differences with limits of agreement

First Author, Publication Year, Country	Study design	Patients Characteristics, Sample Size (n)	Intervention	Comparator(s)	Outcomes
Mangat, 2010,[17] United Kingdom	Non-blinded prospective study. Hospital setting.	61 elective surgical patients scheduled for general anaesthesia. 46 M/15 W. Age: mean 66 years (SD 14).	2 infrared tympanic thermometers (Genius 2 and PRO4000), 1 temporal artery thermometer (Exergen 5000)	Nasopharyngeal temperature probe (Thermistor 400 series 9Fr)	▪ Correlation with reference
Rabbani, 2010,[18] Pakistan	Non-blinded prospective study. Outpatient hospital setting.	2000 patients presenting with or without fever to four departments. 1149 M/851 F. Age: mean 31.8 ± 19.4 years, 626 aged 5-16 years, 730 aged 17-40 years, 478 aged 41-60 years, 166 older than 60 years.	Tympanic thermometer (Beurer FT25)	Oral mercury thermometer	▪ Correlation with reference ▪ Febrile range ▪ Sensitivity ▪ Specificity ▪ PPV ▪ NPV
Dzarr, 2009,[16] Malaysia	Comparative prospective study. Hospital setting.	21 neutropenic cancer patients, with or without fever. 10 M/11 W. Age: range 15 to 63 years old.	Infrared tympanic thermometry (Braun Thermoscan), mercury bulb oral and axillary thermometers	Mercury bulb rectal thermometer	▪ Correlation with reference ▪ Sensitivity ▪ Specificity ▪ PPV ▪ NPV

First Author, Publication Year, Country	Study design	Patients Characteristics, Sample Size (n)	Intervention	Comparator(s)	Outcomes
Smitz, 2009,[19] Belgium	Non-blinded prospective study. In-patient hospital setting.	100 patients admitted to a geriatric unit, with or without fever. 31 M/69 F. Age: mean 80.9 (SD 7.5) years.	2 infrared ear thermometers (ThermoScan PRO 3000 and First-Temp Genius 3000A)	Rectal thermometer (HP 21075A probe and HP 78342A monitor)	▪ Correlation with comparator ▪ Sensitivity ▪ Specificity ▪ PPV ▪ NPV

F = females; ITDS = Infrared thermal detection system; NPV = negative predictive value; M = males; MA = meta-analysis; PPV = positive predictive value; RCT = randomized controlled trial; ROC = receiver operating characteristics; SD = standard deviation; SR = systematic review; USA = United States of America.

APPENDIX 3: Critical Appraisal of Included Studies

First Author, Publication Year, Country	Strengths	Limitations
Systematic Reviews		
Zhen, 2014,[8] China	A priori-designed SR with MAs. Literature search (including grey literature) strategy described and duplicate study selection. The characteristics of the studies are provided along with their quality scores. Heterogeneity and publication bias have been assessed.	A list of excluded studies is not provided, only included studies are reported. Contradictory conclusions were presented in the text. Conflicts of interest were not assessed in the included studies.
Zhen, 2014,[9] China	A priori-designed SR with MAs. Literature search strategy described and duplicate study selection. A description of included studies was reported. Heterogeneity of the data have been assessed.	Excluded studies are not disclosed. Individual quality of studies was assessed, but not reported. Inclusion criteria in some studies were subjective. Publication bias was not properly assessed. Conflicts of interest were not assessed in the included studies.
Jefferies, 2011,[7] New Zealand	A priori-designed SR. Literature search strategy was described. The characteristics of the studies and their quality were reported. Conclusions were in line with the results.	Study selection was not duplicated. The authors did not mentioned if grey literature was included. Excluded studies were not disclosed. Publication bias was not assessed. Conflict of interest were not assessed for included studies. Included studies were heterogenous, lacked some informations on the study or patient characteristics and the statistical methods either failed to account for repeated measures on the same participants or did not report appropriate measures of variation (meta-analysis could not be conducted).
Bitar, 2009,[6] France	The characteristics of the included studies were reported.	An a priori design has not been mentioned. Study selection was not

First Author, Publication Year, Country	Strengths	Limitations
		duplicated. Literature search strategy was not comprehensive, inclusion of grey literature is uncertain. Excluded studies were not properly reported. Quality assessment of included studies was not documented. Publication bias was not assessed. Conflict of interest were not disclosed.
Non-randomized studies		
Cho, 2014,[10] Korea	Description of objective, outcomes, subject characteristics, interventions, findings (with actual *P* value). Participants were representative of the study population and in a realistic context. Comparison of interventions is made on a single group. Outcome measure was objective.	Investigators were not blinded. Retrospective study. Variability of measurements have not been correctly reported. No power calculation. Conclusions based on very small numbers (6 cases of fever). The timing of measurements was not reported.
Chan, 2013,[11] China	Description of objective, outcomes, subject characteristics, findings (with actual *P* value and CI). Participants were representative of a population in hospital. Comparison of interventions is made on a single group. Outcome measure was objective.	Oral and ear thermometers were not described. The percentage of participation was not disclosed. Investigators were not blinded. No power calculation. No description of statistical analyses. The timing of measurements was not reported.
Selent, 2013,[22] USA	Description of objective, outcomes, subject characteristics, interventions, findings (with actual *P* value and CI). A high percentage (80%) of eligible children participated in the study. Comparison of interventions is made on	Investigators and patients were not blinded. The timing of measurements was not reported.

First Author, Publication Year, Country	Strengths	Limitations
	a single group. Outcome measure was objective. Power calculation has been made.	
Chue, 2012,[12] Thailand	Description of objective, participants characteristics, interventions, findings (with actual *P* value and CI). A fair percentage (72.6%) of eligible children participated to the study. Comparison of interventions is made on a single group. Outcome measure was objective. Power calculation has been made. Measurements taken at the same time.	Outcomes were not clearly stated. Very ill patients were excluded from the study. Investigators and patients were not blinded.
Teran, 2012,[23] Bolivia	Description of objective, patient characteristics, interventions, findings (with actual *P* value and CI). Comparison of interventions is made on a single group. Outcome measure was objective. Measurements taken within a short time period.	Outcomes were not clearly described. The percentage of participation was not reported. Investigators and patients were not blinded. Power calculation has not been done.
Chiappini, 2011,[20] Italy	Description of objective, outcomes, subject characteristics, interventions, findings (with actual *P* value) were described. Participants were representative of the study population and in a realistic context. Comparison of interventions is made on a single group. Outcome measure was objective. Power calculation has been made. Measurements taken within a short time period.	Investigators were not blinded. The percentage of participation was not disclosed.

First Author, Publication Year, Country	Strengths	Limitations
Priest 2011,[13] New Zealand	Objective, outcomes, interventions, and findings were well described. Comparison of interventions is made on a single group. Outcome measure was objective. Power calculation has been made and statistical methods well described	Investigators were not blinded. Subject characteristics were not well described. A small proportions of passengers (15.8%) were analyzed. No *P*-values were presented. Conclusions were based on detecting a small proportion of patients with fever (0.5% to 3%). The timing of measurements was not reported.
Barnett, 2011,[14] USA	Objective, patients, outcomes, interventions, and findings were well described. Comparison of interventions is made on a single group. Outcome measure was objective. Measurements taken within a short time period.	Investigators were not blinded. No power calculation was performed. No statistical test have been done. The % of patients who accepted to participate in the study is not disclosed. Whether the participants were representative of the study population is unclear.
Rubia-Rubia 2010,[3] Spain	Description of objective, outcomes, subject characteristics, findings (with actual *P* value and CI) were described. Comparison of interventions is made on a single group. Outcome measures were objective. Power calculation has been made. Measurements taken at the same time.	The authors did not desribe the devices used. The percentage of participation was not reported. Investigators were not blinded.
Fortuna 2010,[21] USA	Description of objective, outcomes, subject characteristics, interventions, findings (with actual *P* value) were described. Comparison of interventions is made on a single group. Outcome measures were objective.	Investigators were not blinded. The percentage of participation was not reported. Power calculation has not been presented.

First Author, Publication Year, Country	Strengths	Limitations
	Measurements taken within a short time period.	
Nguyen, 2010,[24] USA	Description of objective, outcomes, subject characteristics, interventions, findings (with actual *P* value and CI) were described. A high percentage (86%) of eligible patients were enrolled. Comparison of interventions is made on a single group. Outcome measure was objective. Sample size calculation has been made. Measurements taken within a short time period.	Investigators were not blinded.
Oyakhirome, 2010,[15] Gabon	Description of objective, outcomes, interventions were described. Comparison of interventions is made on a single group. Outcome measure was objective. Measurements taken within a short time period.	Investigators were not blinded. Inclusion and exclusion criteria not explicitly stated. Percentage of enrollment was not reported. Gold standard, rectal thermometry, was not measured for children >6 years. Characterisitcs of participants (e.g. proportion of each gender, age range) was lacking.
Mangat 2010,[17] United Kingdom	Objective, outcomes, subject characteristics, interventions, and findings (with *P* value and CI for significant results) were clearly described. Comparison of interventions was made on a single group. Outcome measure was objective. Power calculation and description of statistical methods were provided.	Investigators were not blinded. The percentage of enrollement was not reported. The selection of surgical patients who are presumably afebrile may limit conclusions about device accuracy at higher temperatures. All subjects were warmed by a water mattress during measurements.

First Author, Publication Year, Country	Strengths	Limitations
	Measurements taken within a short time period.	
Rabbani 2010,[18] Pakistan	Description of objective, outcomes, subject characteristics, interventions, findings (with actual *P* value and CI) were provided. All patients were exposed to each intervention. Age subgroups were used to present results. Outcome measure was objective. Measurements taken within a short time period.	Investigators were not blinded. The percentage of enrollment was not reported. No power calculation. Conclusions regarding accuracy for elderly population based on small numbers of febrile cases (n = 7).
Dzarr 2009,[16] Malaysia	Clear descriptions of objective, outcomes, subject characteristics, interventions, findings (with 95% CIs) were provided. Investigators who recorded mercury bulb temperature readings were blinded as to their position on the patient (oral, axillary, rectal). Comparison of interventions was made on a single group. Outcome measure was objective. Power calculation and description of statistical methods were provided. Measurements taken at the same time.	The percentage of enrollement was not reported. Investigators operating and recording infrared tympanic membrane thermometer readings were not blinded. *P* values not provided.
Smitz 2009,[19] Belgium	Descriptions of objective, outcomes, subject characteristics, interventions, findings (with *P* values and CIs) were provided. Participants were representative of a geriatric population in hospital (67% enrollement); severely ill patients were not excluded. Comparison of interventions was made	Investigators were not blinded. No power calculation; sample size was subject to recruitment during a defined study time period.

First Author, Publication Year, Country	Strengths	Limitations
	on a single group. Outcome measure was objective. Measurements taken within a short time period.	

CI = confidence interval; MA = meta-analysis; *P* = probability value; RCT = randomized controlled trial; SR = systematic review; USA = United States of America.

APPENDIX 4: Summary of Study Findings

First Author, Publication Year, Country	Main Study Findings	Authors' Conclusions
Systematic reviews		
Zhen, 2014,[8] China	• Pooled sensitivity was 0.70 (95% CI 0.68 – 0.72) and pooled specificity was 0.86 (95% CI 0.85-0.88). There was high heterogeneity in the pooled results. • The pooled positive likelihood ratio was 9.14 (95% CI 6.37-13.11) and the negative likelihood ratio was 0.24 (95% CI 0.17-0.34). There was high heterogeneity in the pooled results. • The pooled diagnostic odds ratio was 47.3 (95% CI 29.79-75.18). The area under the summary receiver operating characteristic (ROC) curve was 0.94 and the Q* value was 0.87. There was high heterogeneity in the pooled results.	"Contrary to our expectations, our results show that the accuracy of infrared tympanic thermometry in the diagnosis of pediatric fever is moderate, if not poor" (p.6) "Based on our meta-analysis, accuracy of infrared tympanic thermometry in diagnostic fever is high. We can cautiously make the conclusion that infrared tympanic thermometry should be widely used for measuring fever." (p. 7)
Zhen, 2014, [9] China	• The overall pooled mean difference between tympanic and rectal temperature (mercury and electronic) was 0.22ºC (95% limits of agreements -0.44ºC to 1.30ºC). There was significant heterogeneity in the data. • The mean differences between tympanic and subgrouped rectal temperature were: Mercury 0.21ºC (95% LOA -0.44ºC to 1.27ºC) and electronic 0.24ºC (95% LOA -0.46ºC to 1.34ºC). There was significant heterogeneity in the data. • In febrile children (subgroup), the pooled mean difference between tympanic and rectal temperature was 0.15ºC (95% LOA -0.32ºC to 1.10ºC). There was significant heterogeneity in the data.	"The accuracy of infrared ear thermometry in children is poor, and it cannot replace rectal thermometry in clinical practice, especially, for the diagnosis of febrile children" (p. 1163)
Jefferies, 2011,[7] New Zealand	• Five of seven different tympanic thermometer/mode/core temperature range combinations were clinically accurate with a mean difference within ±0.2ºC of core febrile temperatures • The two tympanic thermometer/mode/core temperature	"For the purposes of ongoing clinical practice and trials using oral, tympanic or rectal thermometry, we advise that, on the basis of the limited available evidence, tympanic and oral thermometry methods should be regarded as equivalent to core

First Author, Publication Year, Country	Main Study Findings	Authors' Conclusions
	ranges combinations that exceeded this limit had a mean difference of -0.22ºC (Thermoscan Pro-1/unadjusted mode/ temp. 37.6-38.0ºC) and 0.24ºC (Thermoscan HM-1/oral mode) from core temperature.	temperature and that rectal thermometry should not be used" (p. 199)
Bitar, 2009, [6] France	• Sensitivity varied from 4.0 to 89.6%. • Specificity varied from 75.4 to 99.6%. • The positive predictive values (PPV) varied from 0.9 to 76.0% and the negative predictive value (NPV) from 86.1 to 99.7%. • In 3 studies, reported values of the area under the curves of ROC were of 0.96, 0.92 and 0.86. • Correlation coefficients with the reference (forehead vs tympanic) were of 0.25, 0.51 and 0.71 in 3 studies. • Sensitivity was higher with external auricular meatus vs forehead (compared in 2 studies): 82.7% vs 17.3% and 67.0 vs 4.0%. Specificity remained high: 98.7% and 96.0%. • When fever prevalence was fixed to 1% in all studies, the derived PPV (forehead area) varied from 3.5 to 65.4% and the derived NPV was ≥99%.	"The epidemiological characteristics of the infection play a major role… Sociological factors can also affect the efficacy of border control measures…The psychological reassuring effect on the public can influence the decision to implement such screening,… but public may loose confidence in this measure…policy makers may feel some pressure to use NCIT but the decision making process should not ignore the poor scientific evidence on NCIT's efficacy…" (p.4)
Non-randomized studies		
Cho, 2014,[10] Korea	• Fever prevalence was of 0.002% among the total arrivals screened and of 1% among the symptomatic arrivals. • Among self-reported fever arrivals (31 cases), 2 cases (6.5%) were confirmed as febrile. Of all non-self-reported febrile arrivals (577 cases), 4 cases (0.7%) were identified as febrile. The association with the declaration was statistically significant ($P < 0.001$). • The average temperature from thermal camera scanning (36.83ºC) and average tympanic temperature (38.14ºC) were not statistically different ($P = 0.316$).	"The findings imply that a procedure for mass detection of fever such as self-reported questionnaires and thermal camera scanning may serve as an effective tool for detecting febrile arrivals at quarantine stations." (p.1)
Chan, 2013,[11] China	• 113 (7.4%) of patients had fever. • IRT temperatures were lower (-3.10ºC) and more variable than reference. The	"Infrared thermographic temperature correlates only moderately with core temperature, but performs better in

First Author, Publication Year, Country	Main Study Findings	Authors' Conclusions
	correlation between the two was significant ($P < 0.001$), albeit generally under 0.5. The maximal forehead (FOREMAX) had the lowest correlation. • This correlation was dependent on age, gender and core temperature, i.e. optimal in ≤ 20 years old febrile males. • To detect a core temperature ≥38ºC, the area under the curves of ROC were of 0.812 (95% CI, 0.761-0.863), 0.780 (95% CI, 0.723-0.837), and 0.815 (95% CI, 0.763-0.867) for maximal frontal (AREAMAX), FOREMAX and lateral views (LATMAX) temperatures. No difference when subgrouped in regard to sex. • At a low cut-off temperature (35ºC IRT, i.e. ≈38ºC core temp) for AREAMAX and LATMAX, the maximum sensitivity was 0.87, specificity was 0.34-0.43, PPV was 0.10-0.11, NPV was 0.97-0.98, positive likelihood 1.33-1.53 and negative likelihood 0.29-0.37.	children, men and among febrile subjects… Although the study results suggested better test performances using either the maximum lateral or frontal temperature, their sensitivity might still not be high enough and the high number/proportion of false positives would be overwhelming. This property renders IRT unsuitable as a routine screening tool for febrile conditions, especially at border crossings with huge numbers of passengers. A single IRT measurement of the forehead from a distance should be replaced by a method with greater sensitivity and specificity" (p. 114)
Selent, 2013,[22] USA	• 306 (35.8%) children had confirmed fever. Parents reported fever in 400 (46.8%) children. • At optimal fever threshold, sensitivities for Opto Therm, FLIR and Thermofocus were of 83.0%, 83.7% and 76.8%, respectively. Similar to patient report (83.9%). • Specificity for Opto Therm, FLIR and Thermofocus were of 86.3%, 85.7% and 79.4%, respectively. Higher than parent report (70.8%). • Correlation with traditional thermometry ($P < 0.01$ vs reference) for Opto Therm, FLIR and Thermofocus were of 0.78, 0.75 and 0.66, respectively. • The ROC curves of OptoTherm and FLIR were similar based on ROC contrast tests ($P = 0.8025$), and areas under the curves were similar, 92.2% and 92.3%, respectively. • Thermofocus' area under the curve was significantly lower at 85.2%, and the	"As part of a health care facility's comprehensive infectious disease control plan, ITDSs could be utilized as a noncontact, noninvasive device to objectively screen for fever at a distance protective for health care workers and to reduce the risk of unrecognized infections and subsequent respiratory disease transmission. Infrared thermal detection systems might also be a reasonable solution for mass screening, such as in border settings, within the limits imposed by available resources and as dictated by an outbreak's epidemiology." (p. 310)

First Author, Publication Year, Country	Main Study Findings	Authors' Conclusions
	curve did differ significantly from both OptoTherm ($P < 0.0001$) and FLIR ($P < 0.0001$) based on ROC contrast tests. • Age, antipyretic, use, emotional state and positioning of child with parent in ITDS field were factors affecting readings.	
Chue, 2012,[12] Thailand	• 38.3% of participants were defined as febrile (i.e. ≥37.5ºC with oral). • For each of the three investigators, the mean difference from oral temperature was 0.05 (95% CI, 0.01-0.08)ºC, 0.11 (95% CI, 0.07-0.16)ºC, 0.12 (95% CI, 0.07-0.17)ºC, respectively. • Most (92.0%) differences between tympanic temperature measurements on the same participant were within the manufacturers reported accuracy of ±0.2ºC, and 98.4% within ±0.5ºC. • Ambient temperature affected the difference only slightly ($P = 0.002$), the difference between the two methods of temperature measurement being 0.09 (95% CI, 0.04-0.15)ºC at an ambient temperature of 30ºC and 0.04 (95% CI, -0.01-0.09) at an ambient temperature of 40ºC.	"The tympanic thermometer provides comparable but more rapid results than the oral mercury thermometer even with high ambient temperatures in the rural tropics". (p. 5)
Teran, 2012,[23] Bolivia	• 167 children were identified with fever. • Mean temperature was 37.9 ± 0.9ºC for the rectal mercury thermometer, 37.6 ± 0.8ºC for the temporal artery thermometer and 37.9 ± 0.9ºC for the non-contact infrared thermometer. • The mean difference vs rectal thermometry was of 0.029 ± 0.01ºC for the non-contact infrared and – 0.2 ± 0.277ºC for the temporal artery. • A significant ($P < 0.001$) and strong (0.952 for non-contact infrared and 0.950 for temporal artery) correlation was shown vs rectal temperature. • The sensitivity and specificity of the non-contact infrared thermometer were of 97%. The PPV and NPV were of 95.2% and 98.1%, respectively. • The sensitivity and specificity of the	"The results demonstrated that the non-contact infrared thermometer could be a good option in the measurement of fever in the paediatric emergency room and inpatient unit. The use of this device is especially helpful when there is a high volume of patients and the measurement of temperature is needed frequently and quickly, which is often seen in daily hospital practices…. In conclusion, the non-contact infrared thermometer is a reliable, comfortable and accurate method of measurement of temperature and is a very useful tool to screen for fever in the paediatric population." (p. 475)

First Author, Publication Year, Country	Main Study Findings	Authors' Conclusions
	temporal artery thermometer were of 91% and 99.6%, respectively. The PPV and NPV were of 99.3% and 94.6%, respectively.	
Chiappini, 2011,[20] Italy	• Clinical repeatability was 0.108°C (SD 0.095) for NCIT and 0.114°C (SD 0.103) for mercury-in-glass. • Mean body temperature measured was 37.19°C (SD 0.96) for mercury-in-glass and 37.30°C (SD 0.92) for NCIT (*P* = 0.153). • Using linear regression analysis, a significant correlation was obtained between the two temperature values (r^2 = 0.837; *P* <0.0001). • Diagnostic performance of NCIT in predicting axillary temperature of mercury-in-glass of >38°C by mercury in glass thermometer: o sensitivity = 0.89 (95% CI, 0.80 to 0.97). o specificity = 0.90 (95% CI, 0.86 to 0.94). o PPV = 0.70 (95% CI, 0.590 to 0.81). o NPV = 0.97 (95% CI, 0.94 to 0.99). • The ROC curve to determine best threshold for axillary temperature >38.0°C, for a mid-forehead temperature of 37.98°C the sensitivity was 88.7% and specificity was 89.9%. • Mean distress score was significantly lower for NCIT (*P* <0.0001). • Differences in children's temperature were not significantly correlated to age or room temperature.	"Our data suggest that NCIT may be a good alternative to tympanic infrared devices in children." (p.1316) "According to our results, the NCIT showed good performance in our study population, has the advantage of measuring body temperature in two seconds and is comfortable for children." (p.1316)
Priest 2011,[13] New Zealand	• 0.5% of travelers screened were identified as febrile (temperature ≥37.8°C) using the IT IS. • The area under the ROC curve for ITIS front of face measurement prediction of tympanic temperature ≥37.8°C was 0.86 (95% CI, 0.75 to 0.97) o Sensitivity of 86% gave a specificity of 71% o Prevalence of fever was 0.5% o PPV in this population was 1.5%	"This study shows that, among a group comprising both asymptomatic and symptomatic arriving international airline travellers, ITIS can have moderately high sensitivity and specificity for a high body core temperature of ≥37.8°C. However, the low prevalence of fever in arriving travellers means PPV is very low." (p. 4)

First Author, Publication Year, Country	Main Study Findings	Authors' Conclusions

	• The area under the ROC curve for ITIS front of face prediction of influenza infection with a temperature threshold of 35.4°C was 0.66 (95% CI, 0.56 to 0.75)
	  o Sensitivity of 87% gave a specificity of 39%
	  o PPV of 2.8%
	• None of the 30 travellers who tested positive for influenza had tympanic temperatures ≥37.8°C (95% CI, 0% to 12%) at screening.
		"Our findings therefore suggest that ITIS is unlikely to be effective for entry screening of travellers to detect influenza infection with the intention of preventing entry of the virus into a country." (p. 6)

Barnett, 2011,[14] USA	• The lowest mean temperatures were obtained with oral thermometry and the highest were found via tympanic temperature. • The difference in mean temperature between right tympanic membrane and rectal temperature was 0.22°F (95% CI, -1.61 to 2.05). • Fever (>100.4°F) was identified in 19%, 6% and 25% of patients with rectal, oral and right tympanic membrane thermometry, respectively. • Relative to fever determined by rectal thermometry, the sensitivity, the specificity, the PPV and the NPV of tympanic thermometry were 74.12%, 86.22%, 55.26% and 93.55% respectively. • From ROC curves, the optimal fever cutoff point for tympanic thermometry was determined to be 100.2°F (sensitivity 80.0%, specificity 80.8%, AUC = 0.878). • If the patient's tympanic is between 100.5 and 101.0°F, then the likelihood ratio for a positive test (fever) is 1.74 (95% CI, 0.93-3.26).	"In conclusion, the oral and tympanic temperature readings are not equivalent to rectal thermometry readings. Oral thermometry frequently underestimates the temperature relative to rectal readings, and TM values can either under- or overestimate the rectal temperature. Likelihood ratios help the clinician develop a more precise estimate of a rectal fever based on any given oral or TM reading and alter the posttest probabilities for a rectal fever. When likelihood ratios for a given range of oral or TM readings generate sufficient uncertainly, we recommend that rectal thermometry be used to assess for fever." (p. 511)

Rubia-Rubia 2010,[3] Spain	• Validity for cut-off point pulmonary artery core temperatures 38.5°C, 38.7°C, and 38.9°C
	  o Infrared in right ear (core equivalency)
	  ▪ Area under ROC curve 0.987 ± 0.007, 0.984 ± 0.008, 0.983 ±
		"We consider the tympanic thermometer to be acceptably accurate in core temperature equivalence…With regard to the infra-red frontal thermometer, in our opinion, this device is not very accurate..." (p. 878)

First Author, Publication Year, Country	Main Study Findings	Authors' Conclusions
	0.009 ▪ NPV 98%, 99%, 99% ▪ PPV 89%, 63%, 59% ▪ specificity 98%, 95%, 93% o Infrared in right ear (oral equivalency) ▪ Area under ROC curve 0.967 ± 0.013, 0.960 ± 0.015, 0.972 ± 0.0011 ▪ NPV 98%, 99%, 99% ▪ PPV 64%, 53%, 52% ▪ specificity 91%, 90%, 91% o Infrared frontal on right temple ▪ Area under ROC curve 0.853 ± 0.051, 0.836 ± 0.063, 0.816 ± 0.072 ▪ NPV 96%, 96%, 97% ▪ PPV 47%, 33%, 41% ▪ specificity 83%, 80%, 88%	"If we only evaluate the aspects of validity, reliability, accuracy and external influence, the best thermometer would be the gallium-in-glass for 12 min." (p. 879)
Fortuna 2010,[21] USA	• Average rectal temperature of all participants was 99.6°F (98.7°F to 100.5°). • Average infrared temperature of all participants was 99.5°F (98.6°F to 100.3°F). • Significant monotonic linear relationship between rectal temperatures and infrared thermometry (*P* <0.01) o slope of the regression line was far from unity (0.697 + 0.05, r^2 = 0.48, *P* <0.01). • Infrared thermometry overestimated rectal temperature in patients with lower temperatures. • Infrared thermometry underestimated rectal temperatures in patients with fever (r^2 = 0.149, *P* <0.01).	"Although measurements of surface temperature correlated modestly with rectal temperatures taken contemporaneously, the agreement between the 2 methods was not sufficiently strong to recommend the use of the tested infrared device in clinical practice." (p.103)
Nguyen, 2010,[24] USA	• AUC for OptoTherm was 0.96 (95% CI, 0.94-0.98), FLIR was 0.92 (95% CI, 0.88-0.96), and Wahl was 0.78 (95% CI, 0.72-0.84). • When oral temperature was ≥100°C, OptoTherm identified 275 (11.0%) patients as febrile, sensitivity was 91.0 (95% CI, 85.0-97.0), specificity was 86.0 (95% CI, 81.0-90.0), PPV was	"Our evaluation of 3 [infrared thermal detection systems] in emergency department settings found that the FLIR and OptoTherm reliably identified elevated body temperatures." (p. 1713)

First Author, Publication Year, Country	Main Study Findings	Authors' Conclusions
	17.9 (95% CI, 13.6-22.2), and NPV was 99.6 (95% CI, 99.3-99.8). • When oral temperature was ≥100°C, FLIR identified 247 (9.8%) patients as febrile, sensitivity was 90.0 (95% CI, 84.0-97.0), specificity was 80.0 (95% CI, 76.0-84.0), PPV was 18.4 (95% CI, 13.7-23.0), and NPV was 99.5 (95% CI, 99.1-99.7). • When oral temperature was ≥100°C, Wahl identified 577 (28.0%) patients as febrile, sensitivity was 80.0 (95% CI, 76.0-85.0), specificity was 65.0 (95% CI, 61.0-69.0), PPV was 5.7 (95% CI, 4.1-7.3), and NPV was 99.1 (95% CI, 98.6-99.5).	
Oyakhirome, 2010,[15] Gabon	• Mean difference between rectal ("gold standard") and tympanic was 0.3°C (95% CI, 0.2-0.3) and limits of agreement were -1°C to 2°C. • For a tympanic temperature threshold of 37.5°C: 62% febrile, sensitivity was 81%, specificity was 86%, PPV was 94%, NPV was 65%. • For a tympanic temperature threshold of 38.0°C: 42% febrile, sensitivity was 75%, specificity was 95%, PPV was 94%, NPV was 76%. • For a tympanic temperature threshold of 38.3°C: 35% febrile, sensitivity was 75%, specificity was 93%, PPV was 87%, NPV was 84%. • For a tympanic temperature threshold of 38.7°C: 25% febrile, sensitivity was 68%, specificity was 95%, PPV was 86%, NPV was 87%.	"… electronic rectal measurements were systematically higher than tympanic measurements, which were in turn higher than electronic axillary measurements." (p. 216)
Mangat, 2010,[17] United Kingdom	• PRO4000 IRT demonstrated good agreement with the nasopharyngeal method: mean difference 0.19ºC, 95% limits of agreement –0.32 to 0.71. • Exergen TAT displayed a significant disagreement with nasopharyngeal temperature probes: mean difference 0.66ºC (95% CI, 0.56 to 0.76, *P* < 0.001), 95% limits of agreement –0.15 to 1.48ºC.	"The TAT tested performs variably in febrile subjects (diagnosed with an IRTT). The physiological basis for error with this method of temperature measurement is strong and cannot be ignored. The PRO4000 IRTT gave good performance, and the Genius 2 is best used in its ear mode, where it offers a direct tympanic temperature measurement. The 'oral equivalent'

First Author, Publication Year, Country	Main Study Findings	Authors' Conclusions
	• Genius 2 IRT temperature readings were significantly different from those from the nasopharyngeal probe method: mean difference 0.98ºC (95% CI, 0.91 to 1.05, *P* < 0.001), 95% limits of agreement 0.42 to1.54ºC.	temperature offered by the PRO4000 correlates well with nasopharyngeal (core) temperature, but correcting the measured tympanic temperature to a site distant from the site of actual measurement undoubtedly introduces error." (p. 1114)
Rabbani, 2010,[18] Pakistan	• 97 of 2000 patients (4.85%) were identified with oral mercury temperature (OMT) as febrile. • Mean oral reading was 36.7 ºC (range 34.2 to 40.4, SD 0.66) • Mean tympanic temperature (TT) reading was 36.6 ºC (range 34.1 to 40.0, SD 0.71) • Significant, positive Pearson's correlation (r; *P* < 0.001) between tympanic and oral for all age groups in patients with normal temperatures (overall r = 0.843) • Significant, positive Pearson's correlation (r; *P* < 0.001) between tympanic and oral in febrile patient groups except for those aged 41 and over (age 41 to 60, r = 0.394; age 60 plus, r = 0.452). Overall correlation value between tympanic and oral in febrile patients was 0.723. • Tympanic sensitivity (all age groups) = 66.35 (95% CI, 55 to 75) • Tympanic specificity (all age groups) = 99.63 (95% CI, 99 to 99) • Tympanic PPV (all age groups) = 91.02 (95% CI, 82 to 96) • Tympanic NPV (all age groups) = 98.12 (95% CI, 97 to 98)	"In this study tympanic temperature was found to be a good and quick tool in screening patients, specially [sic] children and young adults, in the OPD setting. This is particularly useful in a high load setting, where rapid temperature measurement decreases patient processing time. There was a poor correlation between TT and OMT in elderly patients. However, the number of elderly patients specially [sic] presenting with fever were very few in this study. A study of an elderly population with a larger number of patients can further explore this correlation." (p. 36)
Dzarr, 2009,[16] Malaysia	• 400 sets of temperature measurements were obtained from 21 patients. • Rectal temperature in 300 randomly selected temperature sets ranged from 35.0 ºC to 41.1 ºC, 66 sets (22%) classified as febrile (≥ 38 ºC). • Intraclass correlation coefficient relative to rectal thermometry was calculated for: o right tympanic (0.810; 95% CI,	"With proper staff training, [tympanic membrane thermometry] is a simple, quick and accurate method of temperature monitoring, which is crucial in the neutropenic setting. We conclude that a single tympanic membrane temperature measurement of either right or left ear is an optimal temperature measuring tool in adult neutropenic patients following

First Author, Publication Year, Country	Main Study Findings	Authors' Conclusions
	0.748 to 0.855) o left tympanic (0.770; 95% CI, 0.713 to 0.815) o mean tympanic (0.806; 95% CI, 0.749 to 0.849) • Sensitivity, specificity, PPV and NPV (in brackets, respectively) of each method to detect rectal fever were as follows: o right tympanic (0.682, 0.979, 0.900, 0.916) o left tympanic (0.712, 0.957, 0.825, 0.922) o mean tympanic (0.636, 0.974, 0.875, 0.905)	chemotherapy. In the absence of contraindication, a corrected oral temperature (+ 0.3 ºC) measurement is a reasonable alternative." (p. 253 to 254)
Smitz, 2009,[19] Belgium	• 18 patients were febrile (RT ≥ 37.8 ºC), 61 patients were afebrile. • Mean rectal temperature in 100 sets of temperature measurements was 37.33 ºC (range 36.3 ºC to 40.7 ºC, SD 0.78). • Mean ear temperature measured with each infrared thermometer was significantly higher ($P < 0.001$) than rectal temperature. • A significant, positive correlation with rectal temperature was shown for the ThermoScan (slope = 0.82; 95% CI, 0.75 to 0.89; $P < 0.001$; r = 0.91) and the Genius thermometer (slope = 0.90; 95% CI, 0.78 to 1.02; $P < 0.001$; r = 0.84) • ThermoScan 95% limits of agreement with rectal temperature were –0.83 ºC and 0.42 ºC (95% CI, –0.88 to 0.48 ºC) • Genius 95% limits of agreement with rectal temperature were –1.32 ºC and 0.20 ºC (95% CI, –1.39 to 0.27 ºC) • Optimal ear fever thresholds were 38.0 ºC (ThermoScan) and 38.3 ºC (Genius) • ThermoScan sensitivity, specificity, PPV and NPV, respectively were 94%, 98%, 89%, and 99%. • Genius sensitivity, specificity, PPV and NPV, respectively were 94%, 96%, 85%, and 99%.	"Infrared ear thermometry can predict rectal temperature within the normal range and in febrile patients, with an acceptable level of accuracy. However, the performance of ear thermometry depends on both operator technique and quality of instrumentation. Additional studies are needed to assess the performance of infrared ear thermometry and to establish recommendations regarding its correct usage." (p. 455)

CI = confidence interval; F = Farenheit; IRT = infrared thermography; ITDS = infrared thermal detection system; ITIS = infrared thermal image scanners; LOA = limits of agreement; NCIT = non-contact infrared thermometer; NPV = negative predictive value; OMT = oral mercury thermometer; OPD = outpatient department; *P* = probability value; PPV = positive predictive value; ROC = receiver operating characteristic; SD = standard deviation; SR = systematic review; TAT = temporal artery thermometer; TT = tympanic membrane temperature; USA = United States of America.